Current Diabetes Reports (2019) 19: 72
https://doi.org/10.1007/s11892-019-1189-3

MICROVASCULAR COMPLICATIONS—RETINOPATHY (DL CHAO AND G YIU, SECTION EDITORS)

Artificial Intelligence Screening for Diabetic Retinopathy: the Real-World Emerging Application

Valentina Bellemo[1] · Gilbert Lim[1,2] · Tyler Hyungtaek Rim[1,3] · Gavin S. W. Tan[1,3] · Carol Y. Cheung[4] · SriniVas Sadda[5] · Ming-guang He[6] · Adnan Tufail[7] · Mong Li Lee[2] · Wynne Hsu[2] · Daniel Shu Wei Ting[1,3]

Published online: 31 July 2019

Abstract

Purpose of Review This paper systematically reviews the recent progress in diabetic retinopathy screening. It provides an integrated overview of the current state of knowledge of emerging techniques using artificial intelligence integration in national screening programs around the world. Existing methodological approaches and research insights are evaluated. An understanding of existing gaps and future directions is created.

Recent Findings Over the past decades, artificial intelligence has emerged into the scientific consciousness with breakthroughs that are sparking increasing interest among computer science and medical communities. Specifically, machine learning and deep learning (a subtype of machine learning) applications of artificial intelligence are spreading into areas that previously were thought to be only the purview of humans, and a number of applications in ophthalmology field have been explored. Multiple studies all around the world have demonstrated that such systems can behave on par with clinical experts with robust diagnostic performance in diabetic retinopathy diagnosis. However, only few tools have been evaluated in clinical prospective studies.

Summary Given the rapid and impressive progress of artificial intelligence technologies, the implementation of deep learning systems into routinely practiced diabetic retinopathy screening could represent a cost-effective alternative to help reduce the incidence of preventable blindness around the world.

Valentina Bellemo and Gilbert Lim contributed equally to this work.

This article is part of the Topical Collection on *Microvascular Complications—Retinopathy*

✉ Daniel Shu Wei Ting
daniel.ting.s.w@singhealth.com.sg

Valentina Bellemo
bellemo.valentina@seri.com.sg

Gilbert Lim
gilbertlim@gmail.com

Tyler Hyungtaek Rim
tyler.rim@snec.com.sg

Gavin S. W. Tan
gavin.tan@singhealth.com.sg

Carol Y. Cheung
carolcheung@cuhk.edu.hk

SriniVas Sadda
vassadda@gmail.com

Ming-guang He
mingguang_he@yahoo.com

Adnan Tufail
Adnan.tufail@moorfields.nhs.uk

Mong Li Lee
leeml@comp.nus.edu.sg

Wynne Hsu
whsu@comp.nus.edu.sg

[1] Singapore National Eye Centre, Singapore Eye Research Institute, 11 Third Hospital Avenue, Singapore 168751, Singapore

[2] School of Computing, National University of Singapore, Singapore, Singapore

[3] Duke-NUS Medical School, Singapore, Singapore

[4] Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong

[5] Doheny Eye Institute, University of California, Los Angeles, CA, USA

[6] Center of Eye Research Australia, Melbourne, Victoria, Australia

[7] Moorfields Eye Hospital & Institute of Ophthalmology, UCL, London, UK

Keywords Artificial intelligence · Deep learning · Diabetic retinopathy screening · Retinal images · Tele-medicine · Survey

Introduction

Diabetic retinopathy (DR) is a microvascular complication of diabetes mellitus, leading to progressive damage to the retina [1]. DR has been deemed as a global public health problem, and diabetes-related visual impairment requires early detection to be prevented. From epidemiology studies, approximately one in three persons with diabetes has signs of DR, and a third of these might suffer from vision-threatening retinopathy, defined as a severe nonproliferative diabetic retinopathy or diabetic maculae edema (DME) [2, 3]. Recently, DR incidence has been estimated to range from 2.2 to 12.7% [4] and remains a leading cause of vision loss in many developed countries, particularly among the adult working-age population. Management and identification of patients at risk for DR condition by risk stratification are crucial [5].

To prevent development and progression of DR condition, regular follow-ups are required and DR screening programs have long been recommended for individuals with diabetes [6, 7]. The international Council of Ophthalmology (ICO) has initiated a comprehensive guide for physicians and ophthalmologists incorporating evidence-based principles with real-world experience with recommendations for diagnosis, definition, screening and referral criteria, follow-up, and management options. However, since the implementation and maintenance of comprehensive and effective programs require substantial resources, DR screening is not widely practiced worldwide as a national program, and many patients with diabetes are unaware of their risk of DR and other related complications [8, 9].

AI Application in Healthcare

Over the past decades, artificial intelligence (AI) has exploded into the scientific consciousness with breakthroughs that are sparkling increasing interests among computer science and medical communities. AI is a branch of computer science in which machines mimic the cognitive function of human mind, making decisions that were regarded to require human cognition. Since the mid-twentieth century, researchers have identified the need for support systems to process the increasing amounts of clinical data required to make clinical decisions. Machine learning (ML) algorithms empower computers that suggest diagnosis or clinical management without direct human intervention, by extracting clinically relevant information from medical data [10].

With recent progress in digitized data acquisition and computing infrastructures, driven by the successful applications of AI and by harnessing the computation power of graphics processing units (GPU), deep learning (DL) has emerged as an attractive ML method within AI [11]. DL was inspired by the operation of the human brain in processing data and involves model architectures called convolutional neural networks (CNN). CNNs operate on the raw data to learn useful features, identifying significant patterns from training datasets [12]. DL has shown substantial potential in medical imaging [13, 14]. These technologies have achieved robust diagnostic performance in detecting major medical conditions from several medical specialties including radiology [15–18], dermatology [19], ophthalmology [20••, 21••], pathology [22], and pediatric diseases [23]. DL-based tool impact is manifold: for clinicians, as image interpretation assistance; for health systems, as workflow improvement; and for patients, in terms of more reliable diagnosis [24].

Although DL-based technologies in medicine are advancing rapidly and multiple retrospective studies have demonstrated that AI can perform on par with clinical experts, most of these tools have not been evaluated in controlled clinical prospective studies [11, 25]. The real-world implementation into patient-care settings is hindered by several practical issues, such as algorithm transparency, records sharing, privacy, and data standardization. In addition, DL algorithms are extremely 'data hungry' and large resources of medical data are required to conduct large-scale studies [26]. However, given the rapid and impressive progress of AI technologies, the medical community is cautiously optimistic that AI might enable higher-capacity and lower-cost care in reality [27–29].

AI Application in Ophthalmology

In the field of ophthalmology, AI using ML and DL has been broadly studied [11, 30]. Over the past few years, several applications in fundus photography, optical coherence tomography (OCT), and visual fields have shown clinically acceptable performance. Specifically, robust classification performance in detection of age-related macular degeneration (AMD) [31–33], glaucoma [34, 35], and retinopathy of prematurity [36, 37] has been achieved from retinal fundus photographs. Promising diagnostic accuracies on par with clinical experts have also been obtained for DR detection [20••, 21••, 38–40].

DL in ocular imaging has been shown to be effective when applied to OCT for AMD automated detection [41–43], macular telangiectasia automated detection [44], visual acuity outcomes [45], and DME-automated segmentation [46] and detection [43]. DL techniques have also been successfully

applied for the detection and quantification of different types of macular fluid from OCT images [47]. In addition, performance comparable to human experts in referral triage decisions and classification of 10 OCT pathologies with DL strategies have been attained [48•]. DL has also been applied to non-traditional task on retinal images such to estimate refractive error [49], quantity of anterior chamber inflammatory cells [50], forecast future Humphrey visual fields [51], and predict cardiovascular risk factors [52–54].

AI Application for Diabetic Retinopathy Screening

One of the most promising areas for AI applications in ophthalmology is in DR screening. For AI to be most effective, it needs a well-run national screening program where all known diabetics are invited to screening at regular intervals. While there have been prior efforts to incorporate automated systems in DR screening workflows and national screening programs, recent advances in DL have seen renewed efforts from many countries to pursue AI to apply to DR screening. Strengths and potential of these studies are clear, as performance indicators have resulted to be better than screening guideline recommendations (typical sensitivity and specificity > 80%) [55]. As diagnostic performance measures, area under the receiver operating characteristic curve (AUC), sensitivity (number of true positive assessments over the number of all positive assessments), and specificity (number of true negative assessments over the number of all negative assessments) are the most commonly adopted.

We cover some recent developments of AI that could potentially be deployed in screening programs from different countries (Table 1).

AI Screening in USA

The USA does not benefit from a national screening program for DR, but many research groups developed their DL system for DR diagnosis. Before DL era, since early 2000s, automated analysis of retinal fundus images raised the interest of Abramoff and collaborators, with methods on the automatic determination of image quality gradeability [56], segmentation of vessels [57, 58], optic disk [59], and lesion detection [60, 61]. In 2008, before the blast of DL era, the team developed a ML system reaching AUC of 0.84, with sensitivity of 0.84 and specificity of 0.64 [62]. Later in 2016 [38], the research team used 1748 retinal images of subjects with diabetes from a publicly available dataset (Messidor-2) to develop their DL system (AlexNet/VGGNet). The International Clinical Classification of Diabetic Retinopathy scale [63] (ICDRSS) was used to classify DR severity. Referable DR was defined as moderate non-proliferative DR, severe non-proliferative DR, proliferative DR, and/or DME; vision-threatening DR was defined as severe non-proliferative DR, proliferative DR, and/or DME. Referable DR detection had AUC, specificity and sensitivity of 97% and 87%, respectively. Vision-threatening DR detection had a sensitivity of 100% and specificity of 91%. The AUCs were above 0.95 and no DME cases were missed.

In April 2018, the United States Food and Drug Administration (US FDA) approved the autonomous DR and DME detection software (IDx-DR) developed by Abramoff el al. for providing diagnostic decisions [39]. Specifically, IDx-DR was reviewed under the FDA's De Novo premarket review pathway, a regulatory process for novel low- to moderate-risk devices without any prior legally marketed device. FDA also granted IDx-DR as Breakthrough Device designation. The software was tested in prospective clinical trials at 10 primary care practice sites throughout the USA. In the pivotal trial, from January 2017 to July 2017, 900 patients with diabetes were enrolled. Wisconsin Fundus Photograph Reading Center specialists graded the fundus photographs and the Early Treatment Diabetic Retinopathy Study Severity Scale [64] (ETDRS) criteria were used for grading. In particular, more than mild DR (mtmDR) was defined as an ETDRS level of 35 or higher and/or DME. The resulting sensitivity of mtmDR was 87% and specificity was 91%.

Other research groups in the USA developed their DL systems for DR detection, using mainly publicly available databases. Gulshan and colleagues [21••] likewise reported excellent diagnostic performances of their DL systems (Inception-V3) in detecting referable DR in 2016. As with Abramoff's team, Google AI adopted ICDRSS with the same definitions of referable DR and vision-threatening DR. They used a large training dataset ($n = 128{,}175$, graded 3 to 7 times for diabetes by a panel of 54 US licensed ophthalmologists and ophthalmology senior residents) and two separate publicly available datasets ($n = 9963$ EyePACS-1 images and $n = 1748$ Messidor-2 images, graded by at least 7 US board-certified ophthalmologists) to test and validate the model. The authors showed that their DL system could achieve sensitivities of above 96% and specificities greater than 93%, with AUC of 0.991. Subsequently, the team conducted a retrospective analysis to demonstrate that an adjudication process, generating a consensus grade from multiple retina specialists, would provide a more robust reference standard for algorithm development and disease detection [65].

Gargeya et al. [40] deployed a customized CNN to identify healthy individuals with no DR from subjects with any DR from publicly available datasets (EyePACS and E-Ophtha datasets). A total of 75,137 photographs

Table 1 Real-world AI screening programs: validation and prospective studies worldwide

Country	DR National Screeening	DL/ML system	Year of publication	CNN	Training data set	Real-world validation data set	Referable DR diagnostic performance	DME diagnostic performance	Grading scale
USA	No	Abramoff et al	2018	AlexNet/VGGNet	Messidor-2	• 10 primary care practice sites from the USA • 892 patients	• 87.2% sensitivity • 90.7% specificity	–	ETDRS
Singapore	Yes, SiDRP	Ting et al	–	Selena (VGGNet)	SiDRP 2010–2013	• 6 primary care clinics in Singapore • Collecting patients	Ongoing	–	ICDRS
UK	Yes, NHS Diabetic Eye Screening Programme	Tufail et al	2016	iGradingM, Retmarker, EyeArt™	-n/a as independent evaluation using real life screening data from regional screening service-	• 1-Region of London, real world screening data • 20 258 patients 2- Further EyeART study in 3 different regional screening programmes prospective on 30 000 patients	•1–93.8% EyeArt v1 sensitivity • 85%. Retmarker sensitivity 2- EyeART v2.1Sensitivity 98.8%	–	NHS Diabetic Eye Screening
India	No	Rajalakshmi et al	2018	EyeArt™ software	–	•1 tertiary care diabetes hospital in Chennai, Remidio Fundus on Phone device •296 patients	• 95.8% sensitivity • 80.2%. specificity	–	ICDRS
		Google AI	–	Inception-V4 (Google AI)	Messidor-2	• 2 eye cares in India • Collecting patients	Ongoing	–	ICDRS
Thailand	Ministry of Public Health program	Raumviboonsuk et al	2019	Inception-V4 (Google AI)	Messidor-2	•hospitals or health centres in 13 health regions in Thailand •7517 patients	• 0.988AUC • 96.9% sensitivity • 95.3%. specificity	• 0.994AUC • 94.0% sensitivity • 98.2%. specificity	ICDRS
China	Lifeline Express Diabetic Retinopathy Screening Program	Li et al	2018	Inception-V3	LabelMe	•36 hospital ophthalmology departments in China •19 900 images	• 0.989AUC(VTDR) • 97.0% sensitivity • 91.4% specificity	–	NHS Diabetic Eye Screening
Australia	NHMRC program	Keel et al	2018	Inception-V3	LabelMe	• 2 urban endocrinology outpatient services in Australia • 96 patients	• 92.3% sensitivity • 93.7% specificity	–	NHS Diabetic Eye Screening
		Kanagasingam et al	2018	Inception-V3	DiaRetDB1, EyePACS,Australian Tele-eye care DR database	•1 primary care in Western Australia •96 patients	• 92% sensitivity • 12% positive predictive value	–	ICDRS
Africa	No	Bellemo et al	2019	VGGNet+ResNet	SiDRP 2010–2013	• Mobile screening unit in 5 urban centres in Zambia	• 0.973 AUC	• 0.942 AUC	ICDRS

Table 1 (continued)

Country	DR National Screening	DL/ML system	Year of publication	CNN	Training data set	Real-world validation data set	Referable DR diagnostic performance	DME diagnostic performance	Grading scale
						• 1574 patients	• 92.25% sensitivity • 89.04%. specificity	• 97.19% sensitivity	
		Hansen et al	2015	Iowa Detection Program (IDP)	Messidor-2	• Nakuru Eye Study in Kenya • 3460 patients	• 0.878AUC • 91.0% sensitivity • 69.9%. specificity	–	NHS Diabetic Eye Screening

AI artificial intelligence, *auc* area under the receiver operating characteristic curve, *DME* diabetic maculae edema, *DR* diabetic retinopathy, *ETDRS* Early Treatment Diabetic Retinopathy Study, *ICDRS* International Cinical Diabetic Retinopathy Disease Severity Scale, *NHS* National Health Service (UK)

from patients with diabetes were analyzed. The model achieved an overall 0.97 AUC with a 94% sensitivity and 98% specificity.

AI Screening in Singapore

Singapore has implemented the Singapore Integrated Diabetic Retinopathy Program (SiDRP) since 2010. SiDRP is a national-level telemedicine-based screening program for DR that involves the centralization of retinal fundus photograph grading from primary care clinics and utilizes trained graders (Fig. 1a). Patients' images are transmitted to an ocular imaging center, where trained graders assess the images for DR severity, typically within a turnaround time of 1 h. The SiDRP telemedicine system has been evaluated to dominate the previous family physician-based system from a cost perspective [66].

In 2017, Ting et al. [20••] developed their DL system (VGGNet-based) to identify referable DR, AMD, and glaucoma using nearly half a million images from a multi-ethnic community. The training set for the DR model consisted of 76,370 images from the *Singapore* Integrated Diabetic Retinopathy Programme cohort (Chinese, Indian and Malay ethnicities). The generalizability of the model has been demonstrated addressing the challenge of considering different patient cohorts (community-, population-, and clinical-based) under different settings and conditions (patients' demographics, glycaemic control, pupil dilation, retinal cameras, width of field, reference standards, graders' qualification). The DL system was tested on 10 external datasets with 112,648 images (Chinese, Indian, Malay, Caucasian, Hispanic, African-American ethnicities). DR levels were defined using the ICDRSS and referable DR was defined as moderate non-proliferative DR or worse, DME, and/or ungradable image; vision-threatening DR was defined as severe non-proliferative DR and proliferative DR. For the internal validation dataset, AUCs of the DL system were 0.936 and 0.958 for referable DR and vision-threatening DR, respectively. Sensitivities and specificities were 91% and 91% for referable DR and 100% and 91% for vision-threatening DR, respectively. For the external validation datasets, the AUCs of referable DR ranged from 0.889 to 0.983 with sensitivities varying from 92 to 100% and specificities from 73 to 92%.

Following the study, the team is planning to integrate the AI algorithm within SiDRP. In the first phase, the AI system will be active alongside the current SiDRP workflow, with clinical leads reviewing whether its performance is acceptable for clinical operation. Assuming acceptable performance, the second phase will involve manual intervention on the AI output results (Fig. 1b), and if all goes well, SiDRP will transition to a third phase where the AI system functions without manual intervention for normal cases (Fig. 1c).

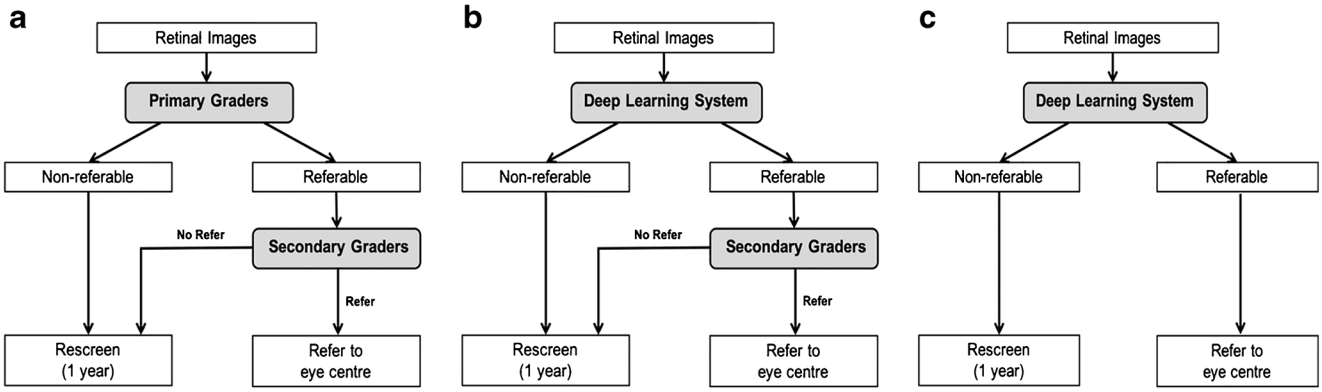

Fig. 1 DR screening models. **a** Manual DR screening model, **b** AI-assisted DR screening model, **c** fully automated DR screening model. Primary and secondary graders are nonmedical professional graders, trained and certified for retinal images grading. According to the International Clinical Classification of Diabetic Retinopathy Severity scale, fundus images are defined as referable with diagnosis of moderate non-proliferative diabetic retinopathy, severe non-proliferative diabetic retinopathy, proliferative diabetic retinopathy, and/or diabetic macular edema. Images classified as ungradable by the dell learning system are considered as referable

AI Screening in the United Kingdom

Screening for DR and vision-threatening DR conditions has been demonstrated to be very effective in England. The National Health Screening (NHS) Diabetic Eye Screening Programme [67] commenced in 2003 and is delivered by qualified clinical and non-clinical staff involved in recognized ongoing Continuous Professional Development and Quality Assurance schemes [68]. The UK National Screening Committee has recommended a systematic population screening program for persons with diabetes aged 12 and over [67] and has achieved a nationwide uptake of 79%.

The first national screening program to deploy (no-DL) image recognition software nationally was in the Scottish DR screening program in 2010 [69–71]. The software was able to assess retinal image quality and detect microaneurysm, local vessels, macular exudates, blot haemorrhages. The Scottish software (iGradingM) was developed in the University of Aberdeen and deployed to evaluate the single 45 degree image per eye approach for screening.

The English DR screening program required at least two 45 degree retinal images per eye and it was felt that before considering commissioning any ML software a properly powered study, independent of developers that evaluated cost effectiveness, should be undertaken [72]. Tufail and coworkers [73] undertook this study that invited automated DR image assessment systems for potential usage in the NHS Diabetic Eye Screening Programme–iGradingM (Medalytix/EMIS Health, Leeds, UK), Retmarker (Retmarker Ltd., Coimbra, Portugal), and EyeArt (EyeNuk Inc., Los Angeles, CA), all ML-based. Retinal images from 20,258 patients attending routine annual DR screening were recruited. When compared with human grader's sensitivity, Retmarker (85%) and EyeArt (94%) systems achieved acceptable performance for referable DR detection and had sufficient sensitivity to make them cost-effectiveness alternatives.

AI Screening in India

In India, 60 million individuals are projected to have diabetes, and 10 to 20% of the diabetic population develops DR. [74, 75] Awareness of this retinal condition is low compared to more developed countries. However, while telemedicine-based DR screening programs are growing and providing valuable clinical benefit [7, 35], the automated interpretation of retinal fundus photographs could be supported using ML and DL tools. The Google AI team is working with eye specialists from Aravind Eye Hospital (Madurai) and Sankara Nethralaya (Chennai) on the development of their AI DL system, with the goal of making such technologies available to everyone.

Rajalakshmi et al. validated a smartphone-based fundus photography system for DR screening [76] and assessed the role of an automated AI algorithm for DR and vision-threatening DR detection [77]. Specifically, images were captured from dilated patients' eyes at a tertiary care diabetes hospital in Chennai, using Remidio Fundus on Phone imaging device, a portable fundus camera. The EyeArt software analysed 296 patients' eyes and provided severity grades according to ICDRSS criteria. The results showed that the AI software exhibited 96% sensitivity and 80% specificity in detecting any DR and 99% sensitivity and 80% specificity in detecting vision-threatening DR.

AI Screening in Thailand

In Thailand, approximately 4.5 million patients with diabetes receive ophthalmic care, but the majority live a great distance from provincial hospitals, where most ophthalmologists and retinal specialists practice. In 2013, the Ministry of Public Health of Thailand set up a national screening program in 13 health regions, with the initial aim to screen in each region at

least 60% of individuals with diabetes. However, data indicates that less than 50% of patients were screened every year [54, 78].

Given the positive results of the Google AI model [65], the team started in December 2018 a partnership with the Rajavithi Hospital in Bangkok, affiliated with the Department of Medical Services, Ministry of Public Health in Thailand. Raumviboonsuk et al. [79] validated the model performance using retinal images from patients with diabetes who participated to the national screening program. The cohort consisted of 7517 patients and 25,326 gradable images. DR and DME were classified according to the ICDRSS guidelines. The reference standard was the grades adjudicated by a panel of international retinal specialists from Thailand, India, and the USA. Referable DR was detected with a sensitivity of 97% by the DL system and 74% by regional graders. The DL system also outperformed humans for DME diagnosis. However, specificities of the DL model were slightly lower than human graders.

AI Screening in China

Diabetes mellitus and prediabetes among adults in China affect 110 million people [80] and a regular screening is extremely challenging to the national healthcare system. Moreover, China faces inadequate practicioner numbers, with only about 20 ophthalmologists per million people (compared to 49 in the UK and 59 in the USA) [81]. Lifeline Express Foundation launched in 2014 the Diabetic Retinopathy Screening Program, which is telemedicine enabled and free. This organization is a multi-hospital program across mainland China, with 30 centers across the country, situated in 15 out of 28 provinces and autonomous regions of China. Since 2016, an AI fundus image processing software (Healgoo) served as a provider for this program. LabelMe (http://www.labelme.org, Guangzhou, China), a labeling platform on which experts can label medical images, was awarded the Medical Device Registration Certificate (Class 2) issued by Guangdong FDA on September 2018 [82, 83].

Recently in 2018, Li et al. [39] conducted a study on a DL-based model (Inception-V3) for the detection of vision-threatening DR, defined as preproliferative DR or worse, and/or DME, according to the NHS scale. The model was trained with 71,043 photographs acquired from LabelMe; 27 ophthalmologists and an experienced ophthalmologist graded the images. The DL system was tested initially on 19,900 LabelMe images with resulting AUC, sensitivity and specificity for vision-threatening DR of 0.989, 97%, and 91%, respectively. Model performance was subsequently evaluated on an independent data set of 35,201 images, including Indigenous Australians from the National Indigenous Eye Health Survey, Malays from the Singapore Malay Eye Study, Australian Caucasians from the Australian Diabetes, Obesity and Lifestyle Study. Overall diagnostic performance for vision-threatening DR detection was 0.955, 93%, and 99% for AUC, sensitivity, and specificity, respectively.

AI Screening in Australia

In Australia, a significant increase in the health and economic impact of DR burden is expected, with the need of enhancements in eye healthcare among Indigenous communities [83, 84]. The National Health and Medical Research Council (NHMRC) has long endorsed DR screening program, recommending annual screening for individuals who have had DR diagnosis, and biennial screening for Australians without significant DR risk factors. However, comprehensive DR screening strategies have not been widely implemented all over the country.

Keel et al. [85] performed a pilot study showing the feasibility of a DL system to screen for DR at two urban endocrinology outpatient services (St. Vincent's Hospital,

Melbourne and University Hospital Geelong, Barwon Health), recruiting 96 adult patients with diabetes between July 2017 and November 2017. The DL model was developed using 66,790 retinal photographs from LabelMe, and the images were graded according to the UK NHS screening guidelines by 21 ophthalmologists. The DL system showed a sensitivity of 92% and a specificity of 94% in referable DR detection. The DL-based automated model was demonstrated to be well accepted by patients, with a 96% of participants that reported to be satisfied or very satisfied.

Also, Kanagasingam et al. [86] in 2018 evaluated the effectiveness of their DL system (Inception-V3) in primary care practice. The model was trained using 30,000 images from publicly available domain (DiaRetDB1, EyePACS, and Australian Tele-eye care DR database) and the severity of DR was based on ICDRSS criteria. A total of 193 patients with diabetes, seen at a primary care in Western Australia, were recruited and 386 images were analyzed. The study reported a resulting specificity of 92%.

AI Screening in Africa

The potential of ML and DL algorithms performing automated retinal images analysis has been demonstrated also on populations from Africa, where radical measures are required to identify and reduce blindness due to diabetes to achieve the Sustainable Development Goals by 2030.

Hansen et al. [87] deployed the ML software of Abramoff team [88] for the detection of sight threatening DR (moderate or more severe DR, and DME), from 6788 fundus photos (3460 participants) of the Nakuru Eye Study, Kenya. The

software performance was compared to grading carried out at Moorfields Reading Centre, according to ICDRSS guidelines. Overall, a sensitivity of 91% and specificity of 70% were achieved.

More recently in 2019, Bellemo et al. [89] demonstrated the potential of AI using DL in a Zambian. The team trained an ensemble DL model (VGGNet and ResNet) on retinal fundus images from patients with diabetes who participated in theSiDRP. The DL system was subsequently validated with retinal images from patients with a self-reported diagnosis of diabetes who attended a mobile screening unit in five urban centers in the Copperbelt province of Zambia. A total of 4504 retinal fundus images from 3093 eyes of 1574 Zambians with diabetes were prospectively recruited. The AUC of the DL system for referable DR was 0.973, with corresponding sensitivity of 92% and specificity of 89%. Vision-threatening DR detection rate was 99% and DME sensitivity was 97%.

Clinical, Technical, and Medicolegal Challenges

Despite considerable recent progress [90] in deep learning-based AI for DR detection as determined by screening performance metrics, many practical challenges remain to be resolved. Notably, cutting-edge expert systems from the 1970s, such as MYCIN [91] and INTERNIST-I [92], have not entered popular practice despite having consistently exhibited performance comparable to clinicians on tasks such as antimicrobial selection [93]. This is largely due to much the same clinician concerns over the interpretability, complexity, and time-effectiveness of AI implementations [94]. With DL in imaging, the main issue has been with the former. In fact, the fundamental nature of pixel-input neural networks has resulted in reduced interpretability as compared to expert systems or even simple logistic regression [95], where the base factors could be identified.

Fortunately, there has been some progress on neural network visualization methods [78, 96] which highlight image regions that were most relevant to the final decision [68, 97]. However, examination of various popular visualization methods such as saliency [82], DeepLift [98], epsilon-LRP [80], Grad*Input [83], and Integrated Gradients [78, 96] reveals that different visualization methods may not always agree with each other (Fig. 2). It is therefore important to recognize the relative strengths and weaknesses of each visualization method.

Given this uncertainty in interpretability, clinician concerns are quite justified, from the threat of legal liability arising from (possibly unreasonably) incorrect AI analyses, concurrent with relatively immature regulatory frameworks for AI systems [99]. On the legal end, current tort mechanisms have been judged to be inadequate for addressing possible injury entailed by AI misdiagnoses [100], despite systems such as

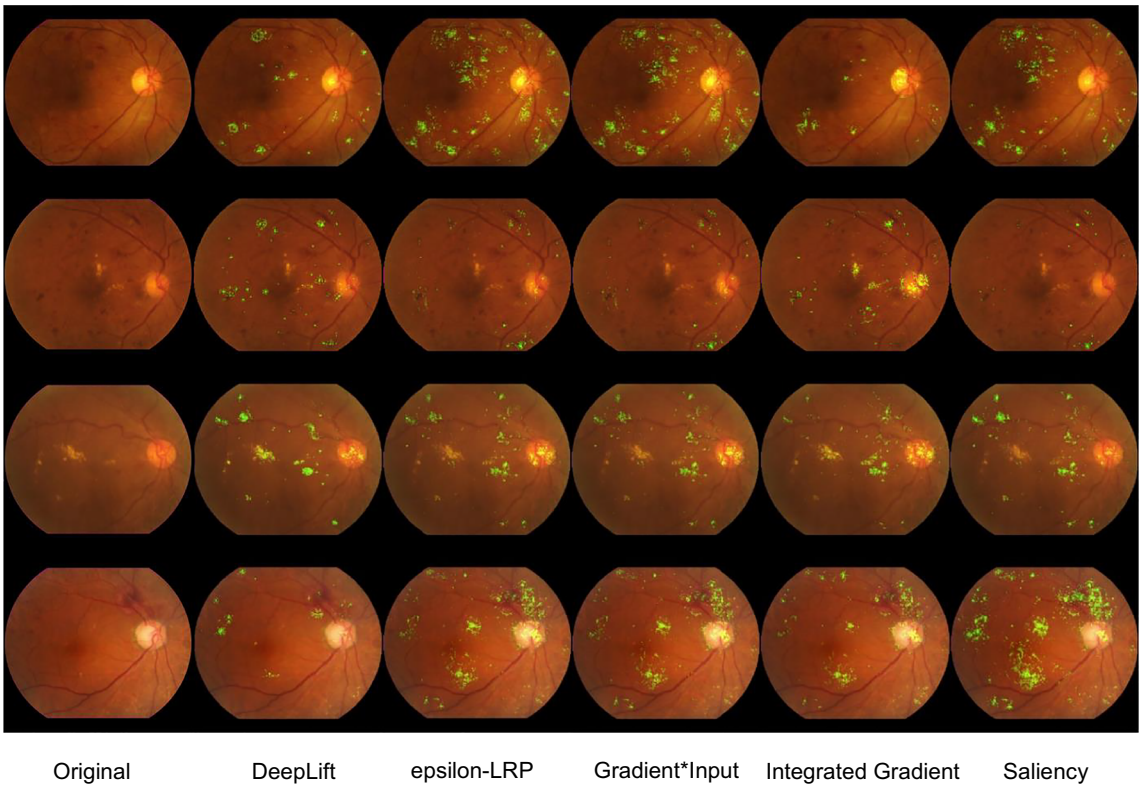

Fig. 2 Heat map visualizations using different models

IDx-DR already having gained FDA approval for DR. [39] Clearly, clinical use of AI will involve some amount of trust on the provider end, with it probably being unreasonable for medical personnel and administrators to attain a deep technical understanding of the AI systems that they are using.

From the other direction, AI researchers will also have to assume that they have been provided with an accurate scope of the task to be achieved, and sufficient resources to create a model for that task. For example, for all the potential of deep neural network models, training good models nevertheless requires many preconditions to be met: the training data should be of sufficient quantity to represent the diagnosis task to begin with, which has tended to involve tens of thousands to millions of images. On top of that, there should be enough examples of individual disease conditions. All these images should also ideally be drawn from realistic environments, and carefully and continually validated over new, independent batches of data.

Future Directions

The key to reducing preventable vision loss worldwide is early detection and prompt treatment of DR. Comprehensive DR screenings are essential. Recently, computer-assisted and automated retinal image assessment tools have been demonstrated to have robust diagnostic performance and can potentially reduce the reliance on manual work and the cost of traditional screening. While such programs have been based on fundus photography, alternative screening platforms using different advanced retinal imaging techniques could be considered (e.g., OCT angiography, adaptive optics, retinal oximetry, metabolic imaging, etc) [101].

With AI progress, especially in imaging domains, there has been a thread of worry about automated systems replacing humans, instead of simply assisting them. The current foreseeable state of AI suggests that a combination of humans and AI will be able to provide the best care, than either alone [102]. This is as they tend to exhibit different, yet complementary, strengths and weaknesses. An AI system is impervious to fatigue and unfailingly detail-oriented, where such details are specified. Humans on the other hand have higher-level metacognition and intuition that allows them to recognize "out-of-set" anomalies more readily.

Additionally, there should be a concurrent evolution of AI algorithms to increasingly take underlying biological mechanisms into account, rather than simply approaching diagnosis as a general ML classification problem. A possible approach would be the decomposition of a classification task into separate intermediate goals [103] conforming to some accepted theory, which would further serve as a sanity check on the model's recommendations, and possibly allow principled hypothesis testing and knowledge discovery to be attempted [104].

Authors' Contributions VB, GL, and DT contributed to the initial drafting of the manuscript. VB, GL, TR, GT, CC, SS, MH, AT, ML, WH, and DT all contributed to the critical review and final approval for this manuscript.

Funding Funding from Research Grants Council-General Research Fund, Hong Kong (Ref: 14102418); National Medical Research Council Health Service Research Grant, Large Collaborative Grant, Ministry of Health, Singapore; the SingHealth Foundation; and the Tanoto Foundation.

Compliance with Ethical Standards

Conflict of Interest Dr. Gilbert Lim, Prof. Mong Li Lee, Prof. Wynne Hsu, and Dr. Daniel S.W. Ting are the co-inventors of the deep learning system for detection of retinal diseases (patent on Automated Retinal Image Analysis Software for Referable Diabetic Retinopathy, Glaucoma Suspect and Age-Related Macular Degeneration 10201706186V [Singapore]). Dr. Ming-guang He is a co-inventor of another deep learning system for detection of retinal diseases (patent on managing color fundus images using deep learning models [ZL201510758675.5]). Valentina Bellemo, Dr. Tyler Hyungtaek Rim, Dr. Gavin S.W. Tan, Dr. Carol Y. Cheung, Dr. SriniVas Sadda, Prof. Ming-guang He, and Prof. Adnan Tufail declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Leasher JL, Bourne RR, Flaxman SR, Jonas JB, Keeffe J, Naidoo K, et al. Global estimates on the number of people blind or visually impaired by diabetic retinopathy: a meta-analysis from 1990 to 2010. Diabetes Care. 2016;39(9):1643–9.
2. YauJW, RogersSL, KawasakiR, LamoureuxEL, KowalskiJW, BekT, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012:DC_111909.
3. Cheung N, Mitchell P. Wong TY. Diabetic retinopathy. 2010;376: 13.
4. SabanayagamC, BanuR, CheeML, LeeR, WangYX, TanG, et al. Incidence and progression of diabetic retinopathy: a systematic review. Lancet Diabetes Endocrinol2018.
5. Ting DSW, Cheung GCM, Wong TY. Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. Clin Exp Ophthalmol. 2016;44(4): 260–77.
6. Klein R, Klein BE. Screening for diabetic retinopathy, revisited. Am J Ophthalmol. 2002;134(2):261–3.
7. Wong TY, Sun J, Kawasaki R, Ruamviboonsuk P, Gupta N, Lansingh VC, et al. Guidelines on diabetic eye care: the

International Council of Ophthalmology Recommendations for screening, follow-up, referral, and treatment based on resource settings. Ophthalmology. 2018;125:1608–22.
8. Wong TY, Cheung CMG, Larsen M, Sharma S, Simó R. Diabetic retinopathy. Nature Reviews Disease Primers. 2016;2:16012.
9. Huang OS, Zheng Y, Tay WT, Chiang PP-C, Lamoureux EL, Wong TY. Lack of awareness of common eye conditions in the community. Ophthalmic Epidemiol. 2013;20(1):52–60.
10. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nature Biomedical Engineering. 2018;2(10):719–31.
11. Lee A, Taylor P, Kalpathy-Cramer J, Tufail A. Machine learning has arrived! Ophthalmology. 2017;124(12):1726–8.
12. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
13. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24–9.
14. FogelAL, KvedarJC. Artificial intelligence powers digital medicine. npj Digital Medicine. 2018;1(1):5.
15. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology. 2017;284(2):574–82.
16. Ting DS, Yi PH, Hui F. Clinical applicability of deep learning system in detecting tuberculosis with chest radiography. Radiology. 2018;286(2):729–31.
17. HwangEJ, ParkS, JinK-N, Kim JI, Choi SY, Lee JH, et al. Development and Validation of a Deep Learning-Based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. Clin Infect Dis. 2018.
18. Titano JJ, Badgeley M, Schefflein J, Pain M, Su A, Cai M, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat Med. 2018;24(9):1337–41.
19. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8.
20.•• Ting DSW, Cheung CY-L, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA. 2017;318(22):2211–23. **Findings from this study suggest that deep learning systems can be generalised on different patient cohorts under different settings and conditions**.
21.•• Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402. **Findings from this study suggest that deep learning can represent a suitable tool with clinical acceptable sensitivity and specificity for detecting referable diabetic retinopathy**.
22. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA. 2017;318(22):2199–210.
23. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25(3):433–8.
24. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
25. Nsoesie EO. Evaluating artificial intelligence applications in clinical settings. JAMA Netw Open. 2018;1(5):e182658.
26. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019;25(1):30–6.
27. Beam AL, Kohane IS. Translating artificial intelligence into clinical care. JAMA. 2016;316(22):2368–9.
28. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317–8.
29. Stead WW. Clinical implications and challenges of artificial intelligence and deep learning. JAMA. 2018;320(11):1107–8.
30. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019;103(2):167–75.
31. Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology. 2018;125(9):1410–20.
32. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA ophthalmology. 2017;135(11):1170–6.
33. Burlina PM, Joshi N, Pacheco KD, Freund DE, Kong J, Bressler NM. Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration. JAMA ophthalmology. 2018;136(12):1359–66.
34. Hood DC, De Moraes CG. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology. 2018;125(8):1207–8.
35. Lim G, Cheng Y, Hsu W, Lee ML, editors. Integrated optic disc and cup segmentation with deep learning. 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI); 2015: IEEE.
36. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. JAMA ophthalmology. 2018;136(7):803–10.
37. TingDSW, WuW-C, TothC. Deep learning for retinopathy of prematurity screening. Br J Ophthalmol. 2018:bjophthalmol-2018-313290.
38. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investigative Opthalmology & Visual Science. 2016;57(13):5200.
39. AbràmoffMD, LavinPT, BirchM, ShahN, FolkJC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. npj Digital Medicine. 2018;1(1).
40. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology. 2017;124(7):962–9.
41. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying Normal versus age-related macular degeneration OCT images. Ophthalmology Retina. 2017;1(4):322–7.
42. Treder M, Lauermann JL, Eter N. Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning. Graefes Arch Clin Exp Ophthalmol. 2018;256(2):259–65.
43. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018;172(5):1122–31.e9.
44. Kihara Y, Heeren TF, Lee CS, Wu Y, Xiao S, Tzaridis S, et al. Estimating Retinal Sensitivity Using Optical Coherence Tomography With Deep-Learning Algorithms in Macular Telangiectasia Type 2. JAMA Netw Open. 2019;2(2):e188029–e.
45. Schmidt-Erfurth U, Bogunovic H, Sadeghipour A, Schlegl T, Langs G, Gerendas BS, et al. Machine learning to analyze the prognostic value of current imaging biomarkers in Neovascular age-related macular degeneration. Ophthalmology Retina. 2018;2(1):24–30.

46. LeeCS, TyringAJ, DeruyterNP, WuY, RokemA, LeeAY. Deep-Learning Based, Automated Segmentation Of Macular Edema In Optical Coherence Tomography. bioRxiv. 2017:135640.
47. Schlegl T, Waldstein SM, Bogunovic H, Endstraßer F, Sadeghipour A, Philip A-M, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning. Ophthalmology. 2018;125(4):549–58.
48.• De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342–50. **Findings from this study suggest that deep learning in making referral recommendations can reach or exceed that of human experts on a range of sight-threatening retinal diseases**.
49. Varadarajan AV, Poplin R, Blumer K, Angermueller C, Ledsam J, Chopra R, et al. Deep learning for predicting refractive error from retinal fundus images. Invest Ophthalmol Vis Sci. 2018;59(7):2861–8.
50. Xiao S, Rokem A, Lee CS, Wilson L, Pepple K, Sabesan R, et al. Fully automated quantification of retinal cones and anterior chamber cells using deep learning. Invest Ophthalmol Vis Sci. 2018;59(9):1222.
51. Wen JC, Lee CS, Keane PA, Xiao S, Wu Y, Rokem A, et al. Forecasting Future Humphrey Visual Fields Using Deep Learning. arXiv preprint arXiv:180404543. 2018.
52. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering. 2018;2(3):158–64.
53. Ting DSW, Wong TY. Eyeing cardiovascular risk factors. Nature Biomedical Engineering. 2018;2(3):140–1.
54. TingDS, CheungCY, NguyenQ, SabanayagamC, LimG, LimZW, et al. Deep learning in estimating prevalence and systemic risk factors for diabetic retinopathy: a multi-ethnic study. npj Digital Medicine. 2019;2(1):24.
55. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. JAMA. 2016;316(22):2366–7.
56. Niemeijer M, Abramoff MD, van Ginneken B. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. Med Image Anal. 2006;10(6):888–98.
57. Niemeijer M, Staal J, van Ginneken B, Loog M, Abramoff MD, editors. Comparative study of retinal vessel segmentation methods on a new publicly available database. Medical imaging 2004: image processing; 2004: International Society for Optics and Photonics.
58. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imaging. 2004;23(4):501–9.
59. AbramoffMD, NiemeijerM, editors. The automatic detection of the optic disc location in retinal images using optic disc location regression. 2006 International Conference of the IEEE Engineering in Medicine and Biology Society; 2006: IEEE.
60. Niemeijer M, Van Ginneken B, Staal J, Suttorp-Schulten MS, Abràmoff MD. Automatic detection of red lesions in digital color fundus photographs. IEEE Trans Med Imaging. 2005;24(5):584–92.
61. Niemeijer M, van Ginneken B, Russell SR, Suttorp-Schulten MS, Abramoff MD. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. Invest Ophthalmol Vis Sci. 2007;48(5):2260–7.
62. Abràmoff MD, Niemeijer M, Suttorp-Schulten MS, Viergever MA, Russell SR, Van Ginneken B. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. Diabetes Care. 2008;31(2):193–8.
63. Wilkinson CP, Ferris FL, Klein RE, Lee PP, Agardh CD, Davis M, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology. 2003;110(9):1677–82.
64. Group ETDRSR. Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12. Ophthalmology. 1991;98(5):823–33.
65. Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology. 2018;125(8):1264–72.
66. Nguyen HV, Tan GSW, Tapp RJ, Mital S, Ting DSW, Wong HT, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. Ophthalmology. 2016;123(12):2571–80.
67. Peto T, Tadros C. Screening for diabetic retinopathy and diabetic macular edema in the United Kingdom. Current diabetes reports. 2012;12(4):338–45.
68. SayresR, TalyA, RahimyE, BlumerK, CozD, Hammel N, et al. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. Ophthalmology. 2018.
69. Scotland GS, McNamee P, Philip S, Fleming AD, Goatman KA, Prescott GJ, et al. Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland. Br J Ophthalmol. 2007;91(11):1518–23.
70. Scotland GS, McNamee P, Fleming AD, Goatman KA, Philip S, Prescott GJ, et al. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. Br J Ophthalmol. 2010;94(6):712–9.
71. Philip S, Fleming AD, Goatman KA, Fonseca S, Mcnamee P, Scotland GS, et al. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. Br J Ophthalmol. 2007;91(11):1512–7.
72. Kapetanakis VV, Rudnicka AR, Liew G, Owen CG, Lee A, Louw V, et al. A study of whether automated diabetic retinopathy image assessment could replace manual grading steps in the English national screening Programme. J Med Screen. 2015;22(3):112–8.
73. Tufail A, Rudisill C, Egan C, Kapetanakis VV, Salas-Vega S, Owen CG, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. Ophthalmology. 2017;124(3):343–51.
74. Raman R, Ganesan S, Pal SS, Kulothungan V, Sharma T. Prevalence and risk factors for diabetic retinopathy in rural India. Sankara Nethralaya Diabetic Retinopathy Epidemiology and Molecular Genetic Study III (SN-DREAMS III), report no 2. BMJ Open Diabetes Res Care. 2014;2(1):e000005.
75. Rema M, Premkumar S, Anitha B, Deepa R, Pradeepa R, Mohan V. Prevalence of diabetic retinopathy in urban India: the Chennai urban rural epidemiology study (CURES) eye study. I Invest Ophthalmol Vis Sci. 2005;46(7):2328–33.
76. Rajalakshmi R, Arulmalar S, Usha M, Prathiba V, Kareemuddin KS, Anjana RM, et al. Validation of smartphone based retinal photography for diabetic retinopathy screening. PLoS One. 2015;10(9):e0138285e.
77. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye. 2018;32(6):1138–44.
78. SundararajanM, TalyA, YanQ, editors. Axiomatic attribution for deep networks. Proceedings of the 34th International Conference on Machine Learning-Volume 70; 2017: JMLR. org.
79. Raumviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, et al. Deep Learning vs. Human Graders for Classifying Severity Levels of Diabetic Retinopathy in a Real-

World Nationwide Screening Program. arXiv preprint arXiv: 181008290. 2018.
80. Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One. 2015;10(7): e0130140.
81. International Council of Ophthalmology : Ophthalmologists Worldwide [updated 2019/04/25/06:38:29. Available from: http://www.icoph.org/ophthalmologists-worldwide.htmlfiles/266/ophthalmologists-worldwide.html.
82. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:13126034. 2013.
83. Shrikumar A, Greenside P, Shcherbina A, Kundaje A. Not just a black box: Learning important features through propagating activation differences. arXiv preprint arXiv:160501713. 2016.
84. Keel S, Wu J, Lee PY, Scheetz J, He M. Visualizing deep learning models for the detection of referable diabetic retinopathy and Glaucoma. JAMA ophthalmology. 2019;137(3):288–92.
85. KeelS, LeePY, ScheetzJ, LiZ, KotowiczMA, MacIsaacRJ, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep. 2018;8(1).
86. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney M-L, Mehrotra A. Evaluation of artificial intelligence–based grading of diabetic retinopathy in primary care. JAMA Netw Open. 2018;1(5):e182665.
87. Hansen MB, Abràmoff MD, Folk JC, Mathenge W, Bastawrous A, Peto T. Results of automated retinal image analysis for detection of diabetic retinopathy from the Nakuru study, Kenya. PloS one. 2015;10(10):e0139148.
88. Abràmoff MD, Reinhardt JM, Russell SR, Folk JC, Mahajan VB, Niemeijer M, et al. Automated early detection of diabetic retinopathy. Ophthalmology. 2010;117(6):1147–54.
89. Bellemo V, Lim ZW, Lim G, Nguyen QD, Xie Y, Yip MYT, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. The Lancet Digital Health. 2019;1(1):e35–44.
90. Naylor CD. On the prospects for a (deep) learning health care system. JAMA. 2018;320(11):1099–100.
91. Shortliffe EH, Davis R, Axline SG, Buchanan BG, Green CC, Cohen SN. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system. Comput Biomed Res. 1975;8(4):303–20.
92. Miller RA, Pople HE Jr, Myers JD. Internist-I, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med. 1982;307(8):468–76.
93. Yu VL, Fagan LM, Wraith SM, Clancey WJ, Scott AC, Hannigan J, et al. Antimicrobial selection by a computer: ablinded evaluation by infectious diseases experts. JAMA. 1979;242(12): 1279–82.
94. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial IntelligenceClinical decision support in the era of artificial intelligence. JAMA. 2018;320(21):2199–200.
95. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N, editors. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015: ACM.
96. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, editors. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision; 2017.
97. Burlina PM, Joshi N, Pacheco KD, Liu TYA, Bressler NM. Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration. JAMA ophthalmology. 2019;137(3):258–64.
98. Shrikumar A, Greenside P, Kundaje A, editors. Learning important features through propagating activation differences. Proceedings of the 34th International Conference on Machine Learning-Volume 70; 2017: JMLR. org.
99. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. Science. 2019;363(6429):810–2.
100. Sullivan HR, Schweikart SJ. Are current tort liability doctrines adequate for addressing injury caused by AI? AMA J Ethics. 2019;21(2):160–6.
101. Goh JKH, Cheung CY, Sim SS, Tan PC, Tan GSW, Wong TY. Retinal imaging techniques for diabetic retinopathy screening. J Diabetes Sci Technol. 2016;10(2):282–94.
102. Liew CJ, Krishnaswamy P, Cheng L, Tan CH, Poh A, Lim T. Artificial intelligence and radiology in Singapore: championing a new age of augmented imaging for unsurpassed patient care. Ann Acad Med Singap. 2019;48:16–24.
103. LimG, LeeML, HsuW. Intermediate goals in deep learning for retinal image analysis. ACCV Workshop on AI for Retinal Image Analysis, ACCV2018, 2019.
104. LimG, LimZW, DejiangX, TingDSW, WongTY, LeeML, et al.Feature isolation for hypothesis testing in retinal imaging: an ischemic stroke prediction case study. Innovative Applications of Artificial Intelligence 2019.